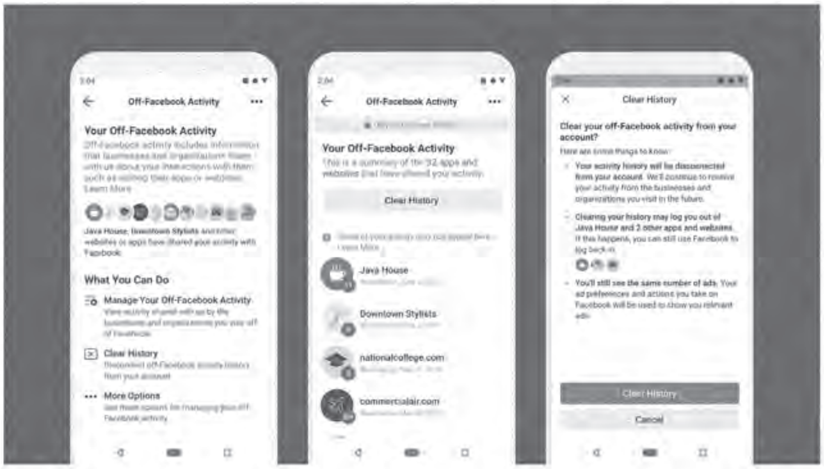
EXHIBIT 52

POSTED ON AUGUST 20, 2019 TO CORE INFRA, DATA INFRASTRUCTURE

Redesigning our systems to provide more control over Off-Facebook activity



Off-Facebook Activity

Your Off-Facebook Activity

Off-Facebook activity includes information that businesses and organizations share with us about your interactions with them, such as visiting their apps or websites. Learn More

Java House, Downtown Stylists and other websites or apps have shared your activity with Facebook

What You Can Do

Manage Your Off-Facebook Activity
View activity shared with us by the businesses and organizations you visit off of Facebook

Clear History
Disconnect off-Facebook activity history from your account

More Options
See more options for managing your off-Facebook activity

Off-Facebook Activity

Your Off-Facebook Activity
This is a summary of the 32 apps and websites that have shared your activity.

Clear History

Java House

Downtown Stylists

nationalcollege.com

commercialair.com

Clear History

Clear your off-Facebook activity from your account?

Here are some things to know:

- Your activity history will be disconnected from your account. We'll continue to receive your activity from the businesses and organizations you visit in the future.
- Clearing your history may log you out of Java House and 2 other apps and websites. If this happens, you can still use Facebook to log back in.
- You'll still see the same number of ads. Your ad preferences and actions you take on Facebook will be used to show you relevant ads.

Clear History

Cancel

By Mayur Patel, Scott Renfro, Benjamin Strahs

At F8 2018, we shared our plan to build a tool that would allow people to see the apps and websites that choose to share activity with us and to disconnect that activity from their accounts. Today we are beginning to roll out that tool, which is called Off-Facebook Activity. Building Off-Facebook Activity required an extensive redesign of the way our systems store and process this activity and new methods for disconnecting information about a person's off-site activity from their account. Since we do not store off-site activity in one centralized profile, we also had to bring together a variety of data sources and develop new methods for managing different types of information across our infrastructure.

To begin developing this tool, we first collected feedback from a variety of groups, including privacy experts, people who use Facebook, and small and large businesses, to understand their needs and ensure we were taking a holistic viewpoint on how to address them. Over the course of these conversations, what emerged was a need for clearer explanations of how data is exchanged on the internet. We also heard that expanding on our controls was a good step in the right direction and aligned with their expectations. We also held a workshop in London with designers, policymakers, academics, and privacy experts, which led to a guide that we hope will serve as a starting point for future work for us as well as the community at large.

We are sharing the story of how we did this, along with some of the choices we made and lessons we learned along the way, with the hope that this may help others interested in building their own data

Related Posts

Related Positions

Software Engineering Manager, Machine Learning
BELLEVUE, US

Software Engineering Manager, Machine Learning
MENLO PARK, US

Software Engineering Manager, Machine Learning
SEATTLE, US

Software Engineering Manager, Machine Learning
NEW YORK, US

Algorithms Software Engineer, Health Technologies - Reality Labs
BURLINGAME, US

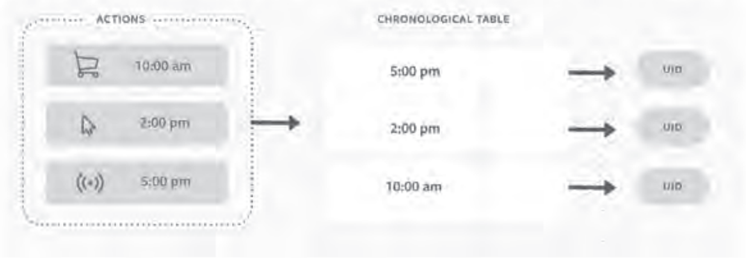
See All Jobs

transparency tools.

Challenge No. 1: Where and how data is stored

As we looked at how to develop this tool, the first task was to assess our existing systems. For content that people share on Facebook — like posts and photos — we use online real-time storage systems that organize information by user and allow people to perform direct deletions. That's how we are able to offer tools like Activity Log. While this type of storage works well for content, the majority of off-site activity information is stored in the data warehouse.

In the data warehouse, we store data in tables and propagate it via pipelines across tables, so that computations can be performed asynchronously. In these tables, information is organized chronologically in daily partitions (similar to a server log) rather than by person. We constantly write new activity data to the end of a table and remove data from the beginning of the table as it expires.



To determine which tables we needed to work with, we ran a discovery process across the data warehouse, starting from the root tables down to all leaf tables. We needed to determine whether the data came from Facebook products (on-site) or from activity and interactions sent to Facebook by businesses and organizations using our business tools (off-site). In order to show people a summary of their activity, we built a centralized copy of key data points that can be queried in real time and is organized by user instead of by time. Once we knew which tables contained activity that would need to be centralized for Off-Facebook Activity, we needed a way to disconnect that information from people's accounts.

Challenge #2: Disconnecting the data

The data warehouse was not designed for deleting or updating individual rows. This allows it to be more efficient at writing, storing, and accessing large volumes of data, but it means that tables cannot be edited on a per-row basis (they can only be replaced wholesale). Since deleting or updating individual rows in real time is not feasible, we had to find a solution that would allow us to modify data in the warehouse for each user, at the tap of a button.

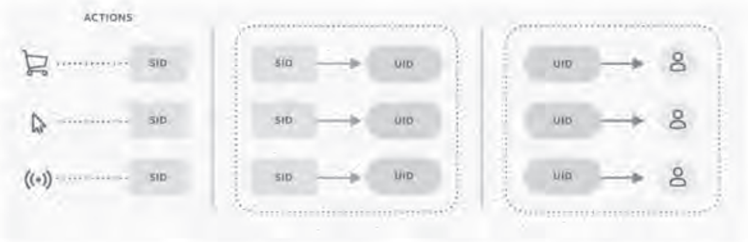
After investigating our options, we made two decisions:

1. Attempting to delete information from various databases across many different tables and rows would take time and may not

work reliably. The quicker, more reliable method would be to disconnect it directly from a person's account.

2. We do not have the capacity to readily remove individual items, and rely instead on bulk methods for disconnecting. Because of this, we decided to offer the ability to disconnect all current off-site activity information in bulk, rather than enabling granular removal of specific activity. We would then enable disconnecting an individual entity's activity going forward.

We implemented these decisions in part by allocating a new separable identifier (SID) for each person, then replacing UID keys with SIDs. We then created a separate mapping between SIDs and UIDs that can be accessed when data is processed. In most cases, when a person uses Off-Facebook Activity to disconnect that information, we remove this mapping between the UID and SID within 48 hours, which breaks the process of joining individual rows in data warehouse tables with a user account. (There are a few special cases, described in more detail below.) We then generate a new SID to be associated with the person's account going forward.



When a person clicks the button to disconnect his or her off-site activity, we remove the mapping between the UID and SID, which breaks the joining of rows in the data warehouse tables to the user account.

If a person chooses to have off-site activity disconnected going forward, we automatically disconnect it and rotate SIDs on a daily basis. We took a similar approach to other information that carries a risk of user identification, like browser or device identifiers, and converted those to alternative SIDs as well. To prevent accidental reassociation between a person's UID and our SIDs, we analyzed tables in the data warehouse to find instances where UIDs and SIDs were stored together. For tables that stored only off-site data and UIDs, we removed user identifiers, replaced user identifiers with SIDs, or removed storage of the table entirely. For tables that contained a mix of both on-site and off-site data, we worked with the owners to split the data into new tables to keep on-site data separate from off-site data.

In some instances, we needed to take different approaches to address specific usage needs, such as when a compound data type containing many separate pieces of information has been stored in a single field and some of those pieces are identifiers. In those cases, we encrypt individual fields using an encrypted ID (EID). The EID functions like a SID, allowing for mapping to the original identifying data. It also sets a time limit on how long this mapping can persist.

When a person clears his or her history, the encryption key is discarded, any mapping from EID to identifying data is made irretrievable, and the data is disassociated from the person's account. This method helps to ensure compliance with our policy that all data must be disconnected within a predefined time frame without requiring a custom SID mapping for every potentially identifying type of data.

Special cases

We also needed to address a few special cases — situations where our primary approach would not work well or where we needed to handle information differently for a specific reason, like guarding against misuse. This involved making decisions about things like what information to show in the Off-Facebook Activity interface, how to handle uses of off-site activity for measurement or security purposes or to guard against fraud, and how to address negative experiences that disconnecting off-site activity might cause, like breaking people's access to apps where they use Facebook Login.

Protecting people's off-site activity

Our original plan was to include in the interface any instance where we retained off-site activity in a way that could be readily connected with a person's account. This includes activity that is associated with a specific device or browser (e.g., via a device ID) but where we cannot authenticate the specific user. For example, we often receive information from browsers that are used by multiple Facebook users or businesses send us information for individuals who visited their site, but may not have been logged in to their Facebook accounts. In situations like these, we were concerned that Off-Facebook Activity might inadvertently show people activity that belongs to other individuals. To mitigate this, we added three protections:

1. We decided to require people to enter a password before accessing the detailed activity list within Off-Facebook Activity, in order to reduce access by anyone else with access to the person's device. This is consistent with protections we provide for other bulk access tools, like Download Your Information and Location History.
2. We decided not to show off-site activity that businesses and organizations sent us in which we could not confirm the person previously used Facebook on that device.
3. We decided not to show off-site activity that businesses and organizations sent us while a person was not logged into Facebook on that particular device.
   Although we do not display this information, we still disconnect it from the person's account.

Measurement

We needed to respect people's desire to disconnect off-site activity from their account and still provide reliable measurement reports to businesses — even if many people choose to disconnect this data from their accounts. To address these dual needs, we created an additional identifier, known as a Measurement ID (MID). Information is tied to the MID, which is connected to a UID only via a separate, access-controlled system. When a person chooses to disconnect his or her off-site activity, we remove the mapping between the MID and the UID and generate a new random MID for that person. This prevents systems from connecting the off-site activity data to that person's UID.

If a person chooses to disconnect off-site activity going forward, we assign him or her a bucketed MID. Unlike MIDs and SIDs, the bucketed MID does not represent an individual user. Instead, it is assigned to many users who have also disconnected off-site activity. With this bucketed MID, we are able to perform aggregated measurement operations — for instance, we can conclude that one of the people in the bucket saw an ad and then visited the target website. We can then aggregate that observation with others who viewed the same ad — without determining exactly which person within the bucket took that action.

Information security and safety

Another important use case is our need to retain some information for security and safety investigations. We wanted to enable people to disconnect their information, but we also needed to ensure malicious actors don't use Off-Facebook Activity to evade enforcement. To solve this, we maintained the ability to flag activity when we see evidence of misconduct, like efforts to access our systems in unauthorized ways or to engage in fraud. In these cases, we can retain a limited set of flagged activity for a longer time. This limited data is stored in separate, access-controlled tables to help ensure that only the relevant security or integrity employees have access to that information. Once the investigation concludes, the data is deleted unless we determine abusive activity has occurred and further action is necessary to protect our products and users.

Policy violations

Our policies articulate which businesses can share information with us and what types of information they can share. For example, we prohibit businesses from sending us health, financial, and other types of sensitive data. Automated systems monitor incoming data to detect certain violations, and humans review the businesses that may be violating our policies. In addition to this monitoring, we also built a feature into Off-Facebook Activity that allows people to provide feedback about the activity they see.

When we identify businesses or activity that violates our policies, we flag the entity that's providing it and do not store any future violating data sent from that entity at time of ingestion (unless we determine that the business has come into compliance and violations are unlikely to occur in the future). We grouped the violating businesses into a section of Off-Facebook Activity that includes entities a person has chosen to disconnect, so that no future data will be associated with his/her account.

Login

Finally, we had to consider how Off-Facebook Activity would interact with Facebook Login. When people disconnect their off-site activity, we need to disconnect information indicating that they previously logged in and used specific apps and websites. This introduced a tension: Disconnecting that data might interfere with the individuals' ability to log in and access their accounts with those third-party apps or websites.

When a person uses Facebook Login to access an app or website, he or she is assigned an app-scoped ID (ASID). This is a stable token that enables the app or website to maintain data associated with the person. Disconnecting the ASID would cause the app or website to not recognize a returning person. To prevent this disruption, we developed a different approach than the standard SID design. We maintain the ASID for people who have cleared their history, but we add noise by storing these ASIDs from connected apps or websites along with hundreds of ASIDs that have never been connected to apps or websites for that person. This allows us to retrieve the stable ASID for third-party apps or websites if a person ever chooses to log back in, but it also prevents us from having a canonical set of apps and websites used by a person who has disconnected their activity.

Upon logging back in to those disconnected apps or websites, a person receives a prompt notifying them that if they choose to continue connecting this app or website, Facebook will enable activity shared by that app or website to be connected with their account. If they enable this feature, they can allow us to store off-site activity for that specific app or website, without having to change their overall setting. If that is not a desired outcome, a person can often choose to create an account that is not powered by Facebook Login instead.

With each of these edge cases, we have carefully thought through the best way to honor people's desire to disconnect their information as well as our commitments to ensure that things like measurement reporting, Facebook Login, and protecting the integrity of our platform continue to work as intended.

While the technical costs for this project were high, we are glad we are able to give people more control over the information that businesses and organizations share with us. This tool is a first not only for us but also for the industry, which means it's very much version 1.

As Mark Zuckerberg explained earlier this year, we are building a privacy-focused messaging and social networking platform. That work includes changing how we think about encryption, data storage, and permanence. We are committed to finding the best path forward and building the technical infrastructure to support this vision. This is going to take time, and there are no easy answers, but we'll keep improving, and we welcome discussions with privacy experts, other companies, and policymakers about how to make controls like this more common across the industry.

Like Share 141 people like this. Be the first of your friends.

◂ Prev
Zoncolan: How Facebook uses static analysis to detect and prevent security issues

Next ▸
Powered by AI: Oculus Insight

Read More in Data Infrastructure

View All ▸

SEP 7, 2023
Arcadia: An end-to-end AI system performance simulator

AUG 29, 2023
Scheduling Jupyter Notebooks at Meta

MAY 16, 2023
Building and deploying MySQL Raft at Meta

MAR 9, 2023
Introducing Velox: An open source unified execution engine

JAN 26, 2023
Tulip: Modernizing Meta's data platform

DEC 12, 2022
Open-sourcing Anonymous Credential Service

Available Positions

Software Engineering Manager, Machine Learning
BELLEVUE, US

Software Engineering Manager, Machine Learning
MENLO PARK, US

Software Engineering Manager, Machine Learning
SEATTLE, US

Software Engineering Manager, Machine Learning

Stay Connected

Engineering at Meta
Like

Meta Open Source
Follow

Open Source

Meta believes in building community through open source technology. Explore our latest projects in Artificial Intelligence, Data Infrastructure, Development Tools, Front End, Languages, Platforms, Security, Virtual Reality, and more.

ANDROID

NEW YORK, US

Algorithms Software Engineer, Health Technologies - Reality Labs

BURLINGAME, US

See All Jobs

Meta Research

Like

Meta for Developers

Like

RSS

Subscribe

iOS

WEB

BACKEND

HARDWARE

Learn More

Meta

Engineering at Meta is a technical news resource for engineers interested in how we solve large-scale technical challenges at Meta.

Home

Company Info

Careers

TermsPrivacyCookiesHelp